Human decision-making

Bounded rationality
Constraints of available time, information, and cognitive capacity limit human ability to make rational decisions

Heuristics and biases
Mental shortcuts that help people make quick, but less accurate decisions, by focusing brain’s limited resources on the most salient information

Human decision-making

Cognitive load

Scarce resources—time, money, food, ... — focus the brain on alleviating shortages and reduce the mental bandwidth to address other needs

“Wisdom of crowds”

Large groups can outperform a few experts on a variety of tasks, e.g., estimating the weight of an ox.

Applications of WoC

- Cultural markets
- Prediction markets
- Crowdsourcing
- Collective deliberation
- Social media
Limits of crowd wisdom

• How does individual’s bounded rationality affect crowd performance?
• What are the impacts of cognitive, or information, overload on crowd performance?
• How can we mitigate the effects of bounded rationality to improve crowd performance?
Outline of the talk

• What are cognitive heuristics and biases?
• How do we measure their impact on individual behaviors and collective outcomes?
  • Controlled experiments
    • MusicLab study [Salganik, Dodds & Watts, 2006]
    • Peer recommendation on Amazon Mechanical Turk [Lerman & Hogg, 2014]
  • Statistical analysis of observational behavioral data
    • Question answering on Stack Exchange [Burghardt et al., 2016]
• What are the potential origins of cognitive biases?
  • Performance deterioration on Reddit [Singer et al., 2016]
Cognitive biases

http://chainsawsuit.com/comic/2014/09/16/on-research/
Examples of cognitive biases

**Position bias:** People pay more attention to items at the top of the screen or a list of items [Payne 1951]

**Social influence bias:** People pay more attention to the popular choices [Buscher et al, CHI’09]
Measuring the effects of cognitive biases

MusicLab study [Salganik et al, 2006]

• Laboratory experiments simulate a cultural market
  • Users presented with a list of songs by unknown artists
  • Users download songs they like $\rightarrow$ popularity
    • Song’s “market share” is the fraction of all downloads it receives

Experimental design

Independent condition (control)

Random order, no download counts

Social influence condition

Ordered by downloads, with download counts

Song quality

Independent condition allows for measuring song quality

Parallel worlds experiments

Repeat the experiment from the same initial conditions and observe the differences in outcome (song market share)

Popularity vs quality

Song market share*: social influence vs independent conditions

* market share = fraction of all downloads in an experiment attributed to a given song
Inequality (A) and unpredictability (B) of song market share in the social influence vs independent conditions worlds.
Conclusions of the MusicLab study

Social influence distorts collective outcomes in markets

- Popularity is only weakly correlated with quality
- Social influence is responsible for the inequality and unpredictability of popularity

But, MusicLab study did not attempt to control for the salience of the social signal – attention paid to songs solely due to their position in the list

Measuring the effects of position bias

Quantify the effects of position bias on individual’s choices and collective outcomes via Amazon Mechanical Turk experiments

No influence condition

Social influence condition

[Lerman & Hogg (2014) “Leveraging position bias to improve peer recommendation” in *PLoSOne*]
Experimental design

- Users asked to recommend stories from a list 100 science stories
  - click on URL to read the story
  - vote for (recommend) the story
- Vary story order → measure popularity (# recommendations)
“Quality”

Fraction of recommendations (market share) in the random ordering

[Lerman & Hogg (2014) “Leveraging position bias to improve peer recommendation” in *PLoSOne*]
Position bias

After accounting for quality, the number of recommendations a story receives in each position gives position bias.

Items in top list positions receive 4x—5x more recommendations than those in lower positions.

[Ormer & Hogg (2014) “Leveraging position bias to improve peer recommendation” in PLoSOne]
Collective outcomes of peer recommendation

Inequality of outcomes: Gini coefficient of the number of recommendations received

Unpredictability of outcomes: Correlation of outcomes across parallel worlds

Position bias results in large inequality and unpredictability when ranking items by popularity

[Lerman & Hogg (2014) “Leveraging position bias to improve peer recommendation” in PLoSOne]
Steering attention improves outcomes

How well does popularity reflect quality?: correlation of popularity with quality

Ranking stories by recency of recommendation results in higher correlation of popularity with quality

[Lerman & Hogg (2014) “Leveraging position bias to improve peer recommendation” in PLoSOne]
Measuring the effects of social influence bias

How does showing social influence signals (popularity) affect individual choices and collective outcomes?

Social influence and collective outcomes

Items with larger social signals get more attention (after controlling for quality and position)

Ratio of actual to expected recommendations

Social signals increase inequality

Gini coefficient

Social influence improves collective efficiency

Social signals reduce effort:
people read and vote for fewer stories with social influence; spend 20% less time

Lessons learned

Cognitive biases—both position bias and social influence—distort collective outcomes of crowdsourcing

• Reduce the correlation between quality and popularity
• Position bias strongly affects individual attention
  • About 2x as important as social influence
• Mitigating the effects of cognitive biases
  • Smarter ordering of items to improve quality of collective outcomes (recency is better than popularity ordering)
  • Judicious use of social signals to reduce individual cognitive effort

Stack Exchange Q&A communities are different. Here's how:

**Expert communities.**
Each of our 161 communities is built by people passionate about a focused topic.

**The right answer. Right on top.**
Experts like you can vote on posts, so the most helpful answers are easy to find.

**Share knowledge. Earn trust.**
Earn reputation and additional privileges for posts others find helpful.
Understanding crowd’s choices
• How do voters choose which answers to vote for?
• How do askers choose which answer to accept?
• How does cognitive load affect these behaviors?

Data
• All user contributions 2009—2014
• 5M questions & answers
• 23M votes
Anatomy of Stack Exchange

Question

What's the correct way to write a `for-in` loop in JavaScript? The browser doesn't issue a complaint about either of the two approaches I show here. First, there is this approach where the iteration variable `x` is explicitly declared:

```javascript
for (var x in set) {
    ...
}
```

And alternatively this approach which reads more naturally but doesn't seem correct to me:

```javascript
for (x in set) {
    ...
}
```

Cognitive load

Number of answers to the question →

Answers

Answer features

- votes/score
- accepted?
- web page order
- chrono order

- num words
- word share
- hyperlinks
- readability
- age

- answerer reputation
- tenure
Statistical analysis of Stack Exchange

Regression model

- Probability to accept an answer or vote for an answer (before or after any answer is accepted)

\[ p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta \cdot x)}} \]

\( x \) = answer features; \( \beta \) = regression coefficients

- Use penalized regression to deal with correlated attributes
Rather than evaluate all answers, people use simple heuristics to choose answers to vote for or accept. Largest coefficients are:

- Web page order → answer’s rank (*cf* position bias)
- Word share → fraction of the screen it occupies (*cf* availability bias)
- Answer acceptance → social proof (*cf* social influence bias)
Cognitive load increases cognitive biases

Regression coefficient for web page order vs cognitive load*

Regression coefficient for word share vs cognitive load*

* using number of answers available to a question as a proxy of cognitive load

→ Cognitive load (information overload) increases reliance on cognitive heuristics
Cognitive load makes behavior more predictable

How well does the model predict new votes and accepts vs cognitive load

→ Behavior becomes more predictable under information overload
  • Heuristics better predict votes and accept decisions in the testing data, as measured by Area Under the Curve (AUC)
  • Votes and acceptance may not be a good measure of quality, especially for popular questions with many answer.
Potential origins of position bias

Why are people less likely to pay attention to information in lower positions on a page?

- **Limited time, growing boredom**
  - Decreasing utility to continue the task leads to task abandonment
- **Directed attention fatigue**
  - Competition from distracting stimuli exhausts the brain & makes it more difficult to maintain attention
- **Ego depletion** [Baumeister et al, 2007]
  - Self-control depends on finite cognitive resources (eg, glucose in the brain). Exerting self-control depletes the resources, creating fatigue and impairing subsequent self-control
Potential origins of position bias

Why are people less likely to pay attention to information in lower positions on a page?

- **Limited time, growing boredom**
  - Decreasing utility to continue the task leads to task abandonment
- **Directed attention fatigue**
  - Competition from distracting stimuli exhausts the brain & makes it more difficult to maintain attention
- **Ego depletion** [Baumeister et al, 2007]
  - Self-control depends on finite cognitive resources (eg, glucose in the brain). Exerting self-control depletes the resources, creating fatigue and impairing subsequent self-control

- **Cognitive depletion hypothesis**: performing an activity online is associated with a deterioration in performance; the longer the activity, the greater the deterioration
  - Investigate by mining behavioral data
Is there short-term deterioration in user performance on Reddit?

Data

- Comments posted by Reddit users in April 2015
  - 40 million comments
  - 2.7 million Reddit users

Challenges in mining behavioral data

- Behavioral data is heterogeneous
  - Composed of groups of different size
  - ... and with different behaviors
- Simpson’s paradox
  - A trend exists in different groups but disappears or reverses when groups are aggregated
- Kidney stones example

<table>
<thead>
<tr>
<th></th>
<th>Treatment A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small stones</td>
<td>Group 1 93% (81/87)</td>
<td>Group 2 87% (234/270)</td>
</tr>
<tr>
<td>Large stones</td>
<td>Group 3 73% (192/263)</td>
<td>Group 4 69% (55/80)</td>
</tr>
<tr>
<td>Both</td>
<td>78% (273/350)</td>
<td>83% (289/350)</td>
</tr>
</tbody>
</table>
Simpson’s paradox in data science

Exposure response in social media: Probability of response (using a hashtag) after K exposures by friends

Exposure response in social media: Probability of response (retweet) after K exposures by friends, disaggregated by num. friends

Romero et al. (2011) in WWW-2011

Hodas & Lerman (2012) in SocialCom
Simpson’s paradox in data science

Slowdown: Facebook users appear to spend more time reading each story over the course of a session.

Speedup: When the same data is disaggregated by length of the session, users spend less reading each story.
User sessions

Control for (some of) the heterogeneity by identifying user sessions on Reddit

- Segment user commenting activity into sessions
  - Session = period of activity without a 60 min or longer break

- Randomized session data set
  - Shuffle time intervals to randomize user activity
  - Check for trends in the shuffled data – if they disappear, proper trends were identified
Measuring performance

We measure user performance by the quality of the comments posted. Comment quality is measured by:

- **Text length**
  - Number of characters in a comment
- **Score**
  - Number of votes the comment receives
- **Number of responses**
  - Number of responses the comment receives
- **Readability**
  - How easily comment can be understood, measured by the Flesch reading ease formula
Later comments become shorter and less popular

Later comments receive fewer responses and become less complex

![Graph showing the relationship between comment position and average responses and readability.](chart)

11% average average responses

13% average average readability

Trends disappear in randomized data

(a) Original session data

(b) Randomized session data

➢ The more time people spend online, the worse they perform
Conclusion: Big data = big opportunities

Availability of large-scale behavioral data has vastly expanded opportunities for discovery in the cognitive and behavioral sciences

- Evidence for bounded rationality in online behaviors
  - Rather than evaluate all available information and choices, people rely on simple cognitive heuristics

- Impact of cognitive heuristics on user choices
  - People rely on simple cognitive heuristics to make decisions, especially as their cognitive load increases
  - This reduces the quality of collective performance, e.g., in crowd computation tasks
  - Exerting mental effort degrades performance over time
THANK YOU!

Questions?
@KristinaLerman
lerman@isi.edu
Human decision-making

Cognitive load
Scarce resources—time, money, food, ... — focus the brain on alleviating shortages and reduce the mental bandwidth to address other needs.

Information overload
Individual capacity to process information can’t keep pace with growing demands resulting in cognitive dysfunctions, mental fatigue, & stress.

Cognitive heuristics and limited attention

Brain focuses its limited resources on salient information

- Cf. visual attention
  - Brain processes small fraction of visual field in real time
  - Size, motion, and contrast make information salient, directing scarce cognitive resources to it
  - Rising cognitive load shrinks useful visual field
Conclusions of the MusicLab study

• Popularity is only *weakly correlated* with quality
• **Social influence** is responsible for the inequality and unpredictability of popularity

“we also note that inequality increased when the salience of the social information signal was increased from experiment 1 to experiment 2. Thus our results suggest … that as individuals are subject to stronger forms of social influence, the collective outcomes will become increasingly unequal.”

Social influence improves collective efficiency

Social influence reduces effort: people read and vote for fewer stories with social influence; spend 20% less time

... but does not degrade outcome: no significant change in correlation of popularity with quality

Cognitive load increases reliance on heuristics

Regression coefficient for web page order vs cognitive load*

... for word share vs cognitive load

... for answer eventually accepted vs cognitive load

* using number of answers available to the question as a proxy of cognitive load
How do you know if the trend is real?

Check the trend in randomized data

Example: does it take longer to make a bigger purchase

[Kooti et al. (2016) “Portrait of an Online Shopper”, in WSDM-2016]
Trend within a subgroup

Subgroup = people who made 5 purchases

![Graph showing trend in average normalized price over days from last purchase for normal and shuffled groups.]

- **Normal**
- **Shuffled**
User sessions

- Segment user commenting activity into sessions
  - Session = period of activity without a 60 min or longer break
- Randomized session data set
  - Shuffle time intervals to randomize user activity

Original session data

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>10 min</td>
<td>7 min</td>
<td>210 min</td>
<td>2 min</td>
<td>11 min</td>
<td></td>
</tr>
</tbody>
</table>

Randomized session data

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>2 min</td>
<td>210 min</td>
<td>11 min</td>
<td>7 min</td>
<td>10 min</td>
<td></td>
</tr>
</tbody>
</table>
Standardized test prep

• Students taking practice tests online
  • 2.8 attempts by
  • 180 thousand users to answer
  • 6 thousand questions

• Performance during test sessions
  • Session is a period of activity without a break (breaks are self-imposed)
  • Performance is measured relative to user’s ability and question difficulty
    • $P = 0 \rightarrow$ user performing as expected
    • $P < 0 \rightarrow$ user is under-performing
    • $P > 0 \rightarrow$ user is performing better than expected
Performance during a session

Performance declines approaching a break: users become less accurate than expected

... but users answer questions faster: they are guessing?

Learning declines: users less likely to remember correct answers to questions they see closer to break

Performance improves following a break: longer breaks result in higher accuracy
Cognitive depletion model

- Findings are consistent with cognitive depletion: answering questions consumes finite cognitive resources, which impairs performance
  - *cf “Ego depletion”* [Baumeister et al, 2007; 2008]
    - Exerting mental effort impairs executive function and self-control
    - Linked to glucose depletion
- Kinetic model of cognitive resource depletion
  - One-resource model: glucose model
    - Primary resource consumed during mental work, recovers during breaks
    - Performance depends on levels of primary resource
  - Two-resource model: lactate shuttle (LASH)
    - Primary resource (lactate), responsible for performance, but normally low
    - Mental work consumes primary resource, and causes conversion of secondary resource (glycogen) to primary resource
    - Secondary resource recovers during breaks
Comparison of models

- Two resource model better explains performance than the one resource model
- Explains over 10% of uncertainty in performance
  - Changes 50:50 odds of answering a question correctly to 70:30 odds
  - Improve performance by better managing resources: e.g., breaks

| Mutual Information of Performance Measures and Estimated Cognitive Resources |
|-------------------------------|-----------------|-----------------|-----------------|
|                                | Two-Resource Model | One-Resource Model |
| Quantity                      | Bits  | Std. Dev. | % of Entropy  | Bits  | Std. Dev. | % of Entropy  |
| MI(A;R)                       | 0.12  | 0.015     | 12%            | 0.04  | 0.006     | 4%            |
| MI(L;R)                       | 0.10  | 0.02      | 16%            | 0.03  | 0.014     | 3%            |
| MI(T;R_b)                     | 0.51  | 0.05      | 8%             | 0.15  | 0.01      | 2%            |
| MI(ΔT;R)                      | 1.18  | 0.06      | 16%            | 0.27  | 0.016     | 4%            |

Table 1: Abbreviations used are MI (mutual information), A (answered correctly), R (resources at beginning and end of question), L (learned correctly), T (time to next question), ΔT (time answering), R_b (resources at beginning of question).